Measurement results: For each device, the *gm_IV/gm_RF* comparison is performed at three different points of the characteristics. These three points correspond to low, medium and high values of V_{gs} and a constant value of V_{ds}. At each point, *gm_IV* is calculated from pulsed IV curves and *gmIRF* is extracted from the pulsed S-parameter measurements. Then, the relative discrepancy is calculated at each point as

$$Discrepancy(\%) = \frac{gm\_RF - gm\_IV}{gm\_IV} \times 100$$

Fig. 1 shows typical results obtained for a PHEMT from the first foundry characterised with the HP pulse system. Note that at V_{gs} = –0.7V, the discrepancy is < 10%. The discrepancy increases as V_{gs} gets higher and finally we obtain a value of *gm_RF* that is almost 40% higher than *gm_IV* at V_{gs} = 0V. Fig. 2 presents typical

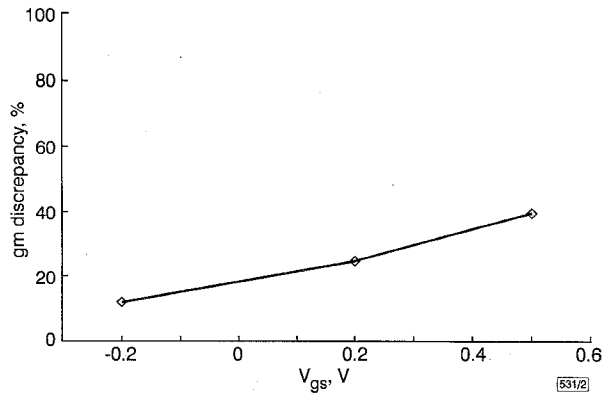

Fig. 2 *gm discrepancy obtained for PHEMT from second foundry, measured with IRCOM pulse system*

V_{ds} = 3V

results obtained for a PHEMT from the second foundry characterised by using the IRCOM pulse system. Again, we can notice that the discrepancy is small at low V_{gs} voltages but increases until a value of *gm_RF* that is 40% higher than *gm_IV* at V_{gs} = 0.5V. The results concerning the HEMT device are shown in Fig. 3. This device has been measured in both pulsed systems. We can again observe a *gm_RF* that is significantly larger (by 50%) than *gm_IV* at high V_{gs} values (V_{gs} = 0.8V). Notice that both pulse systems provide similar results. The same analysis has been performed for different values of the drain voltage, giving the same behaviour in the *gm* comparison.

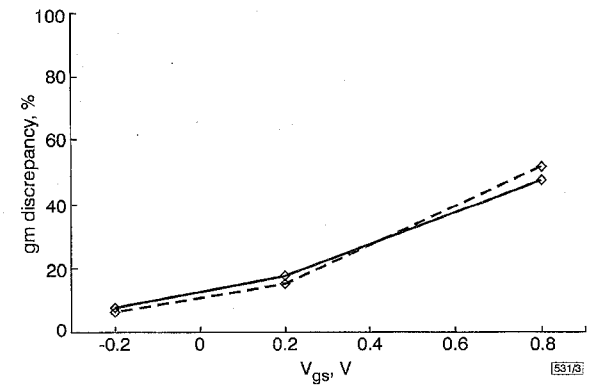

Fig. 3 *gm discrepancy obtained for HEMT device, measured with both pulse systems*

——— IRCOM pulse system
– – – – HP pulse system
V_{ds} = 3V

Discussion and conclusions: Because thermal effects are avoided in this analysis, it is thought that the discrepancies obtained in the comparison of *gm_RF/gm_IV* originated from changes in the trap state during the pulsed I-V characterisation. This means that trap time constants exist that are smaller than the pulsewidth used for the I-V characterisation (600ns in our case); therefore, they can react to the change of V_{gs}. Under such conditions, the trap state is no longer fixed by the quiescent bias point but changes with each pulse level and the pulsed characterisation is no longer performed under isotrapping conditions. Since the study shows a *gm_IV* more than 40% smaller than *gm_RF* at high voltages, nonlinear models extracted from pulsed I-V characteristics can lead to inaccurate results in large-signal simulations.

Discrepancies between *gm_RF* and *gm_IV* in HEMT and PHEMT devices have already been presented [6] using DC and CW data. However, to our knowledge, this is the first time that pulsed systems have been used to perform the comparison, which eliminates self-heating as the origin of the discrepancy and reveals the existence of trap time constants faster than 600ns.

17 November 1997
Electronics Letters Online No: 19980197

J.M. Collantes (*Electricity and Electronics Department, University of Pais Vasco, 48080 Bilbao, Spain*)

Z. Ouarch and R. Quere (*IRCOM, University of Limoges, 7 rue Jules Valles, 19100 Brive, France*)

C.-Y. Chi and M. Sayed (*Hewlett Packard Company, 1400 Fountaingrove Parkway, Santa Rosa, CA 95403, USA*)

References

1 CAMACHO-PENALOSA, , and AITCHINSON, C.S.: 'Modelling frequency dependence of output impedance of a microwave MESFET at low frequencies', *Electron. Lett.*, 1985, **21**, (12), pp. 528–529
2 LADBROOKE, , and BLIGHT, S.R.: 'Low-field, low-frequency dispersion of transconductance in GaAs MESFET's with implications for other rate-dependent anomalies', *IEEE Trans. Electron Devices*, 1988, **35**, pp. 257–267
3 TEYSSIER, J.P., VIAUD, J.P., RAOUX, J.J., and QUERE, R.: 'Fully integrated non-linear modeling and characterization system of microwave transistors with on-wafer pulsed measurements'. IEEE MTTs Digest, Orlando, 1995, pp. 1033–1036
4 COLLANTES, J.M., BARATAUD, D., RAOUX, J.J., and QUERE, R.: 'Complete non-linear MOSFET model for radiomobile applications'. INMMC'96, Duisburg, Germany, October 1996, pp. 102–107
5 SCOTT, J., PARKER, A., RATHMELL, J., and SAYED, M.: 'New applications for pulsed/isothermal test systems'. ARFTG, San Francisco, CA, June 1996, pp. 70–75
6 ANHOLT, R.: 'Electrical and thermal characterization of MESFETs, HEMTs and HBTs' (Artech House 1995), pp. 166–169

Energy function for learning invariance in multilayer perceptron

HanChuan Peng, Lifeng Sha, Qiang Gan and Yu Wei

A new energy function is proposed for forming self-adapting ordered representations of input samples in a multilayer perceptron. Simulation results on unconstrained handwritten digit recognition give a better invariance extraction for this model than for several other models.

Energy function: The mechanism of the visual system for extracting invariances from significantly varying input patterns has been discussed for a long time, but the McCulloch-Pitts neuron has not been accepted as providing a reasonable explanation for the origination of such an ability as associating clearly unlike patterns, unless a more complicated neuron model or a network of simple units is employed [1]. Recently, a trace neural network [2, 3] has been proposed for learning self-adapting representative codes of input samples. Unfortunately, that network cannot give a satisfying performance for realistic tasks. Here, we propose a new energy function for gaining a sparse trace neural network which gives an improved invariance extraction performance.

Consider an M-class classification problem treated by a three-layer perceptron, which contains a linear input layer (IL) with I neurons, a nonlinear hidden layer (HL) with J neurons for feature finding, and a nonlinear output layer (OL) with M neurons for classification. The HL consists of sigmoid units with outputs defined as $y_j(t) = f[\Sigma_{i=1}^{I} x_i(t)w_{ji}(t)]$, where $f(z) = 1/(1+e^{-z})$ is the activation function, x_i is the ith input, and w_{ji} is the connecting weight from the ith input unit to the jth unit in the HL. The nonlinearity

of neurons in the OL is the same as that in the HL. We extend Földiák's definition for the trace of a neuron [1] to the trace of the mth class of input patterns, $T_m(t) = \eta T_m(t-1) + (1-\eta)y(t)$, where η is the trace factor, and its component $T_{mj}(t)$ is the trace of the jth unit $y_j(t)$. Then, we constrain neurons in the HL to minimise the following self-energy function (SEF):

$$B(t) = \sum_{m=1}^{M} B_m(t) = \frac{1}{2}\sum_{m=1}^{M}\sum_{j=1}^{J}[T_{mj}(t-1) - y_j(t)]^2 \quad (1)$$

Here, $B_m(t)$ stands for the sum of the self-energy of all neurons in the HL when the mth class of pattern is presented at a time t. Clearly, the minimisation of $B(t)$ will result in small discrepancies between the outputs of the HL and the related traces. Hence, features produced by the HL can be relatively unchanged but will also be affected by variant samples. $B(t)$ is very different from Stone's model [4] which uses the quotient of a long-term average on the neuron's activation and a corresponding short-term average to achieve a similar effect. Notice that the slow convergence of the minimisation of $B(t)$ and its zero-solution are prevented by summing the square of the difference between the current output and the previous trace (but not the current trace).

To some degree, the trace can be observed as the intra-class average. It is helpful for the post-processing when relatively steady traces are learned. However, learning such an 'average' on a large training set will result in overlapping traces, corresponding different classes, and thus poor post-processing performance. Notice that $B(t)$ only describes the behaviour of particular neurons, but the overlap of traces is indeed the behaviour of the neuron cluster HL. One way to attack this contradiction is to define a mutual energy function (MEF) to make use of the mutual influence of neurons. Generally, an even function which monotonously increases in the positive semi-axis can be a candidate of the MEF, e.g. $S(y) = 0.5\Sigma_{j=1}^{J} \ln[1+y_j^2(t)]$. Actually, by minimising such a function the number of neurons with the same activation can be reduced and, in many cases, some sensible sparse codes of input patterns can be learned [5]. These sparse codes are somewhat factorial with reduced statistic dependence between each other and are advantageous for sequential processing. We define the total energy of the invariance extractor HL to be the weighted sum of SEF and MEF:

$$E = SEF + r \cdot MEF = B + rS \qquad (r \geq 0) \quad (2)$$

where r is a sparseness factor. In this case, the multi-layer perceptron can be called the 'sparse trace neural network' (STNN).

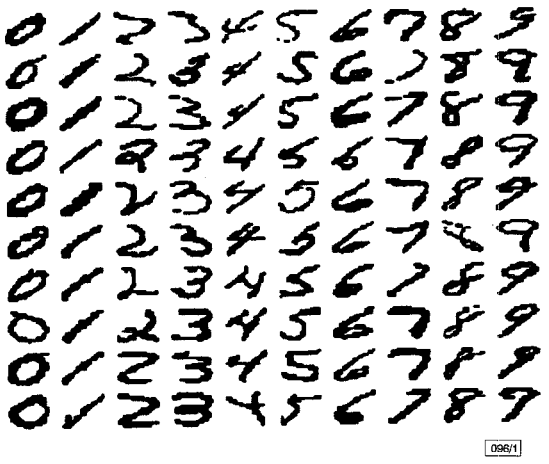

Fig. 1 *Typical samples in digit database*

Algorithm and simulation: We apply the following gradient-descent learning scheme to minimise E:

$$\Delta w_{ji}(t) = -\alpha\frac{\partial E}{\partial w_{ji}(t)}$$

$$= \alpha\left[T_{mj}(t-1) - y_j(t) - r\frac{y_j(t)}{1+y_j^2(t)}\right][1-y_j(t)]y_j(t)x_i(t) \quad (3)$$

where α is a learning factor. Weights can be prevented from being too large by the constraint $\Sigma_{i=1}^{I} w_{ji}^2 = c$, where the constant c is the amplitude of normalisation. The feature extractor HL self-organises according to eqn. 3 to find invariances, which are used by the backpropagation (BP) type classifier OL for categorising input patterns.

A typical example for examining the ability of invariance extraction is unconstrained handwritten character recognition [2, 3]. An unconstrained handwritten digit database from CENPARMI lab in Canada is used to investigate the feature extraction ability of STNN. This database contains 6000 digits in which we randomly selected 4800 digits as a training set while using the other 1200 as a testing set. Typical samples of the handwritten digits are shown in Fig. 1. All samples are normalised to 16×16. 5×5 sized weight masks are used to connect the IL and the HL (a 5×5 block in the IL is connected by a weight mask to a neuron in the HL). Therefore, the IL, HL and OL are 20×20, 12×12 and 10, respectively, with empty margins of two pixels around input patterns. In our experiments, all weights are initialised to be random numbers from a uniform distribution on [–1,1]. Initial traces are obtained by averaging the feedforward computed results of all training samples. All parameters are chosen to be as close as possible to those employed by Wallis [2, 3].

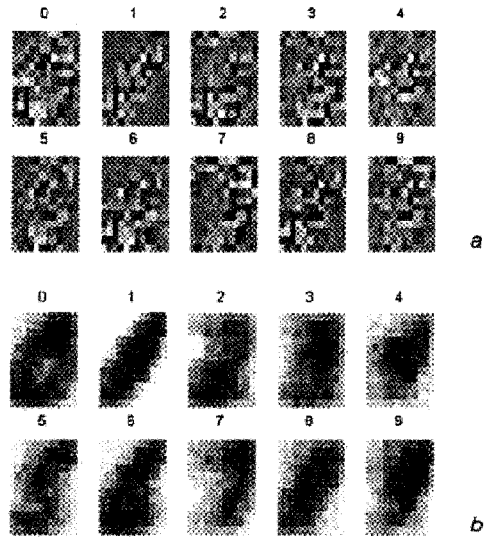

Fig. 2 *Traces before and after learning*

a Initial traces
b Final traces

A small part of the training set is first used to examine the invariance extraction from short sample sequences [2]. 100 samples are picked out randomly in the whole training set (10 samples per class) for training. The average recognition rate of the STNN on the whole testing set is shown in Table 1, where the recognition rate needed in training is also given in parentheses. For comparison, a recent result on a similar database given by Wallis is listed.

Table 1: Recognition rates on testing set

Net model	Short sample sequences	Long sample sequences
	%	%
Wallis [2]	55 (95)	—
BP net	57.92 (95)	90.5 (>99)
STNN	64.83 (95)	90.5 (≈91)

Also, results of the standard BP network with an identical topology are shown. It is quite clear that the invariance extraction layer in the STNN is stronger than that in the models used for comparison. In fact, Wallis's model is far from giving a satisfactory performance for realistic character recognition tasks. However, even when the whole training set is used to investigate the invariance extraction from long sample sequences, the STNN still performs remarkably for its close recognition rates on the training set and the testing set, which implies that the STNN can extract the most interesting invariance for characterising input patterns. On the contrary, the larger difference for the BP network indicates that its

trained weights are not as good at characterising the testing set. The outstanding ability of the STNN for feature finding can be further noted by comparing the initial and final traces, as shown in Fig. 2. The traces can evolve from the initial disorder (Fig. 2*a*) to obvious ordered structures (Fig. 2*b*). This property cannot be observed when only $B(t)$ is employed in learning. Hence, the importance of MEF can be validated as helping the self-clustering of classes.

Discussion and conclusion: Our STNN scheme is quite different from the optimal training signal mapping model presented by Wallis [3]. The STNN does not only show good invariance extraction on a small training set (short pattern sequences), as accomplished by Wallis's model, but also exhibits a good learning ability for a large training set in realistic tasks (long pattern sequences). For this purpose, a new energy function for self-organising the neural network feature extractor is presented. Extracting the invariance from input patterns is a self-organising minimisation process of such an energy function. Our simulation has shown inspiring, good results.

Acknowledgment: We would like to thank G. Wallis for providing some experimental data. This work is supported by a grant from NSF of China (No. 69501003) and a grant from Young Scientists Foundation of Jiangsu Province, China (No. BQ94003).

21 October 1997
Electronics Letters Online No: 19980161

HanChuan Peng, Lifeng Sha, Qiang Gan and Yu Wei (*Department of Biomedical Engineering, Southeast University, Nanjing 210096, People's Republic of China*)

E-mail: phc@seu.edu.cn

References

1 FÖLDIÁK, P.: 'Learning invariance from transformation sequences', *Neural Comput.*, 1991, **3**, pp. 194–200
2 WALLIS, G.: 'Using spatio-temporal correlations to learn invariant object recognition', *Neural Netw.*, 1996, **9**, (9), pp. 1513–1519
3 WALLIS, G., and BADDELEY, R.: 'Optimal unsupervised learning in invariant object recognition', *Neural Comput.*, 1997, **9**, (4), pp. 883–894
4 STONE, J., and BRAY, A.: 'A learning rule for extracting spatio-temporal invariances', *Netw.*, 1995, **6**, pp. 429–436
5 OLSHAUSEN, B., and FIELD, D.: 'Emergence of simple-cell receptive field properties by learning a sparse code for natural images', *Nature*, 1996, **381**, pp. 607–609

Bidirectional transmission of 40Gbit/s WDM signal over 100km dispersion shifted fibre

Chang-Hee Lee, Sang-Soo Lee and Seo Yeon Park

The authors demonstrate bidirectional transmission of a 40Gbit/s (4 × 10Gbit/s) WDM signal over 100 km of dispersion shifted fibre at ITU-T standard wavelengths. The cross-talk penalties induced by Rayleigh back-scattering and four-wave mixing are suppressed by using an arrayed-waveguide grating demultiplexer and channel allocation, respectively.

Introduction: Bidirectional transmission over a single fibre has been investigated for efficient use of optical fibre and for bidirectional optical networks. In bidirectional transmission systems, Rayleigh back-scattering and optical reflection degrade the system performance and induce an optical power penalty [1, 2]. The induced penalty was suppressed by using a fibre grating, and bidirectional transmission of a 10Gbit/s signal over 240km of DSF (dispersion shifted fibre) was demonstrated [3]. For a WDM signal, bidirectional transmission at 40Gbit/s (4 × 10Gbit/s) over 100km of DSF was reported [4]. Two amplification bands of an EDFA were used, i.e. 1530 and 1550nm bands, to reduce the induced penalty. In this Letter, we demonstrate bidirectional transmission of a 40Gbit/s (4 × 10Gbit/s) WDM signal over 100km DSF at ITU-T standard wavelengths in the 1550nm amplification band. The cross-talk induced by Rayleigh back-scattering and optical reflection is suppressed by using an AWG (arrayed-waveguide grating) demultiplexer. The four-wave mixing penalty is also minimised by channel allocation.

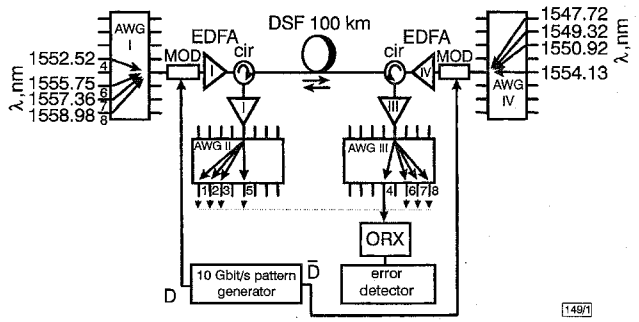

Fig. 1 *Experimental setup for bidirectional tranmission of 4 × 10Gbit/s WDM signal over 100km DSF*

AWG: arrayed-waveguide grating; cir: circulator; MOD: modulator; ORX: optical receiver

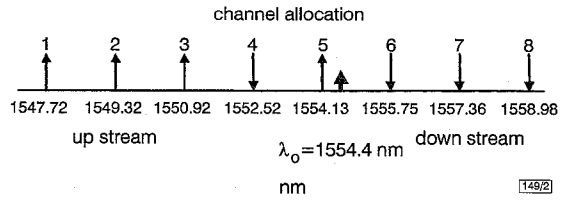

Fig. 2 *Channel allocation diagram*

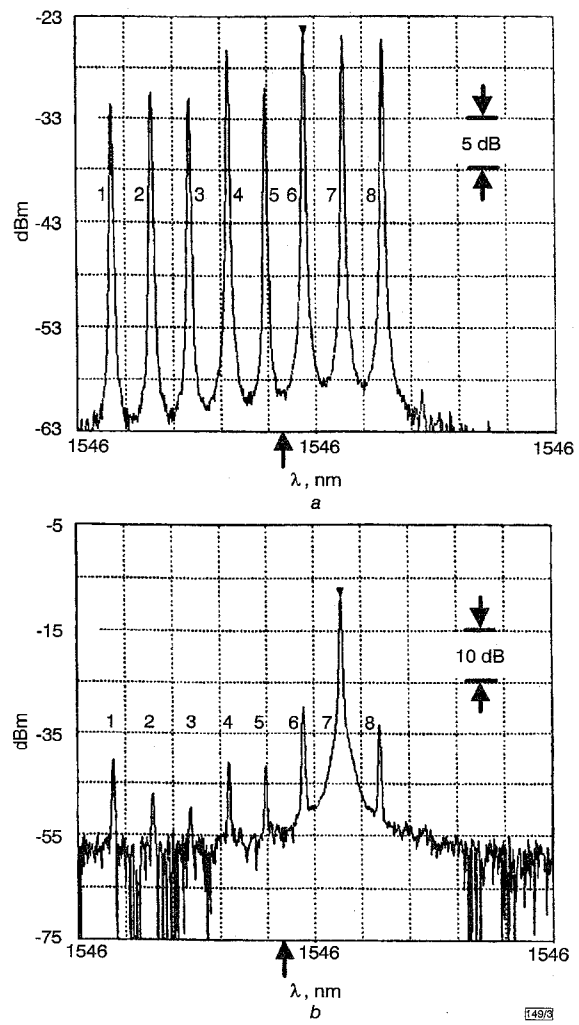

Fig. 3 *Measured optical spectra*

a Input spectrum of EDFA III
b Demultiplexed output spectrum for channel 7